t-Test für abhängige Stichproben

Inhaltsverzeichnis

t-Test für abhängige Stichproben

Lernhinweise

Mit dem t-Test für abhängige Stichproben können Sie entscheiden, ob sich die Mittelwerte von zwei abhängigen Stichproben so stark unterscheiden, dass Sie bei der Interpretation der Differenz den Zufall ausschliessen können.

- Sie können entscheiden, wann der t-Test für abhängige Stichproben eingesetzt wird.
- Sie können in einem einfachen Beispiel den t-Test für abhängige Stichproben "manuell" anwenden.
- Sie können anhand des t-Tests für abhängige Stichproben entscheiden, ob sich die arithmetischen Mittelwerte aus zwei abhängigen Stichproben zufällig unterscheiden.

Benötigte Vorkenntnisse

Sie können diesen Lernschritt effizient bearbeiten, wenn Sie die in Ihrem Curriculum vorgesehene Vorbereitungsliteratur bearbeitet haben und mit dem Grundprinzip der entscheidungsstatistischen Verfahren vertraut sind.

Lernziele

Durch die Bearbeitung des Lernschritts können Sie folgende Fähigkeiten erwerben:

Hinweise zur Bearbeitung des Lernschritts

- Empfohlene Bearbeitung: Beginnen Sie mit "Einführung". Klicken Sie auf den Button "next"; er führt Sie auf den empfohlenen Pfad.
- Wenn Sie auf eine Rubrik klicken, gelangen Sie automatisch zur entsprechenden ersten Seite. In der Regel gelangen Sie zum Inhaltsverzeichnis der angewählten Rubrik.
- Für die Bearbeitung sollten Sie ca. eine Stunde Lernzeit einplanen.

Hinweise zur Bearbeitung des Fallbeispiels

- Das Fallbeispiel ist so konstruiert, dass die Aufgaben problemlos mit einem Taschenrechner (mit Wurzelfunktion) gelöst werden können. Sie brauchen kein Statistikprogramm wie z. B. SPSS.
- Das Ziel der manuellen Bearbeitung ist, dass Sie das Grundprinzip des t-Tests für abhängige Stichproben handelnd nachvollziehen können.
- Es kann sein, dass sich Ihre Resultate aufgrund von Rundungsfehlern in den Dezimalstellen von den Musterlösungen leicht unterscheiden. Das darf Sie nicht verwirren. Das Fallbeispiel ist so konzipiert, dass die Entscheidungen zur Signifikanz eindeutig ausfallen.

Der vorliegende Lernschritt zum **t-Test für abhängige Stichproben** bildet zusammen mit den Lernschritten **F-Test** und **t-Test für unabhängige Stichproben** eine Einheit. Das Fallbeispiel berücksichtigt alle drei Testverfahren.

Einführung

Wenn Sie die Lernschritte zum F-Test und zum t-Test für unabhängige Stichproben nicht kennen, so erfahren Sie die Ausgangslage des Fallbeispiels im Zeitungsartikel

Lokalanzeiger von Gerlechingen

Jahrgang 120 Auflage 20 000 Erscheint monatlich Juli 2002
Preis Fr. 1.25 Inserateannahmeschluss jeweils am 15. des Vormonats

Doping am Schuelerwett–kampf?
Seit ueber achtzig Jahren wird in unserer Gemeinde ein Schuelerwettkampf der Mittel–stufenklassen durchgefuehrt. In diesem Jahr wurde der Anlass durch den Verdacht auf einen Doping–Fall ueberschattet.

Seit über achtzig Jahren berichten wir auch von diesem herrlichen Anlass der Mittelstufenklassen unserer Gemeinde. Das Motto dieses Jahres hiess: "Auch fit im Kopf". Die 105 Kinder wurden durch Losentscheid in zwei Gruppen aufgeteilt. Die eine Gruppe nannte sich "Auerhahn" und die andere "Bibihenne". Neben den sieben sportlichen Aufgaben, hatten die Kinder einen Test zur Merkfähigkeit zu lösen. Die Gruppe "Bibihenne" oder auch Gruppe B genannt, gewann in dieser Disziplin. Ein erfreulicher Erfolg. Doch der Schein trügt, die Enttäuschung war riesengross. Der Hauswart fand im Aufenthaltsraum der Gruppe "Bibihenne" einige Getränkedosen mit der Aufschrift "Stimubrain". Ein Getränk, das die Denk- und Merkfähigkeit von Kindern steigern soll. Es soll auch wirksam gegen Schulmüdigkeit sein.
Waren die Kinder der Gruppe "Bibihenne" gedopt?
Die Lehrerinnen und Lehrer wandten sich an die Schulpflege und baten sie um Hilfe bei der Klärung des Falls.

Frau Fischer von der Schulpflege hat in der Schulpflegesitzung vom 15. Juli angekündigt, dass sie weitere Abklärungen treffen werde, um die unerfreuliche Sache mit dem "Dopingvorwurf" gegen die Gruppe B (Bibihenne) zu klären.
Bevor Frau Fischer ihr Vorgehen zur Lösung des Fallbeispiels vorlegt, traf sie vorgängig Abklärungen beim Kantonalen Labor, bei der Kantonalen Bildungsdirektion, bei den Eltern, bei den Kindern, bei der Lehrerschaft und bei der Produktionsfirma von "Stimubrain".

1. Vom Kantonalen Labor wollte sie wissen, ob das Getränk "Stimubrain" Substanzen enthält, die als gesundheitsschädlich einzustufen sind.
2. Von der Bildungsdirektion wollte sie wissen, ob sie Kindern "Stimubrain" ausschenken dürfe, ohne gegen ethische und rechtliche Richtlinien zu verstossen.
3. Von den Eltern wollte sie wissen, ob sie einwilligen würden, wenn sie die Kinder für einen Test einsetzen würde.
4. Von den Kindern der Gruppe A wollte sie wissen, ob sie bei einem Versuch mit "Stimubrain" mitmachen würden.

5. Von der Lehrerschaft wollte sie wissen, ob die Kinder der Gruppe A für einen Nachmittag frei bekämen.
6. Von der Produktionsfirma von "Stimubrain" wollte sie wissen, ob sie dafür garantiere, dass das Getränk nicht gesundheitsschädlich sei, und ob sie der Auffassung sei, dass das Getränk die Gehirnaktivitäten fördere.

Das Untersuchungsdesign von Frau Fischer wird in der Rubrik "Fallbeispiel" vorgestellt.

Theorie (1-7)

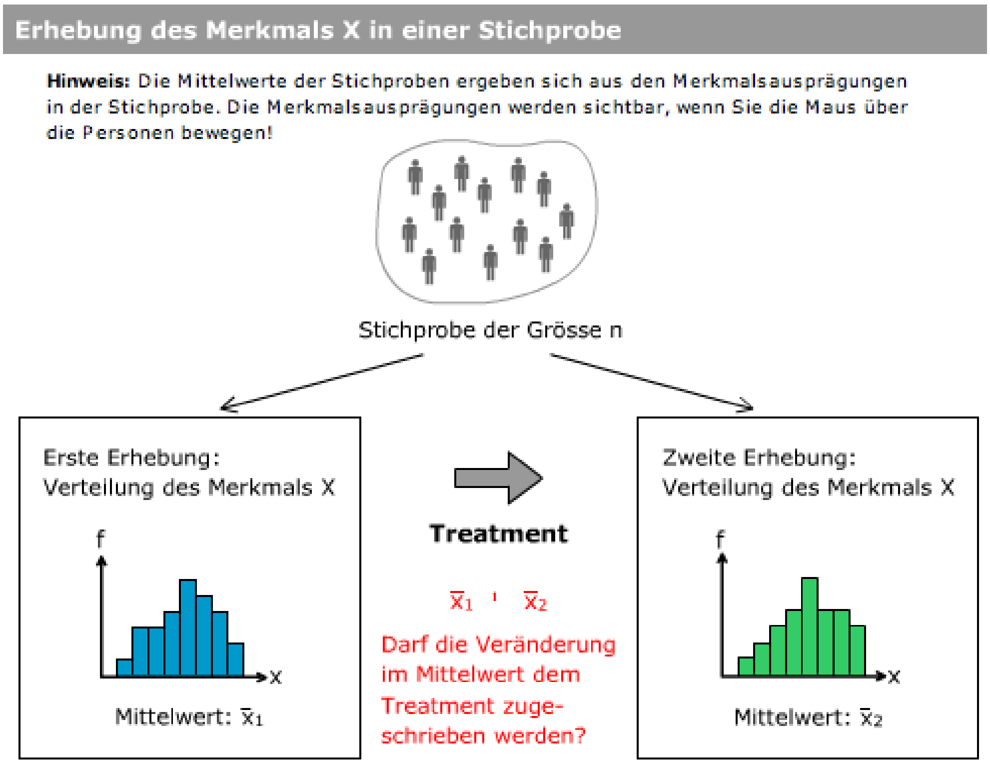
Inhaltsübersicht

1. Inhaltliche Fragestellung

Eine Stichprobe von n Probandinnen und Probanden wurde zweimal untersucht. Das wichtigste und auch häufigste Beispiel, das im folgenden besprochen werden soll, betrifft ein Untersuchungsdesign mit Messwiederholung. In diesem Fall wird in derselben Stichprobe ein Merkmal X zweimal erhoben, einmal vor und einmal nach einem Treatment. (Man spricht auch von einer Vorher- und Nachher-Messung.)

Als Treatment bezeichnet man dabei jede Art von ‚Behandlung' der Stichprobe, die zwischen den beiden Datenerhebungen stattgefunden hat. So ist im Rahmen einer Untersuchung zur Wirksamkeit einer Kundenwerbung die Werbekampagne das Treatment. Im Rahmen einer klinischen Wirkungsforschung hingegen ist das Treatment die zwischen den beiden Erhebungszeitpunkten applizierte Therapie.

Wir widmen uns im folgenden dem Fall, dass das interessierende Merkmal X intervall- oder proportional skaliert ist und damit die Mittelwerte der Merkmalsausprägungen in den beiden Erhebungen bestimmt werden können. Dabei wird sich in der Regel zeigen, dass die beiden Mittelwerte nicht identisch sind.



Da man sich in Untersuchungsdesigns mit Messwiederholung für die Auswirkung des Treatments auf das erhobene Merkmal X interessiert, muss geklärt werden, ob der Unterschied zwischen den beiden Mittelwerten so gravierend ist, dass bei der Interpretation des Unterschieds der Zufall ausgeschlossen werden kann.

2. Statistisch bearbeitbare Fragestellung

Bei der Suche nach einer statistisch bearbeitbaren Formulierung unserer Frage helfen uns die folgenden Überlegungen:

- Unser Interesse gilt den Veränderungen in den Merkmalsausprägungen zwischen der ersten und der zweiten Erhebung. Da dieselben Personen zweimal untersucht wurden, liegen von jeder Person i zwei Merkmalsausprägungen vor; eine aus der ersten Erhebung

  $$x_{i1}$$

  und eine aus der zweiten Erhebung

  $$x_{i2}$$

  , wir sprechen von gepaarten Daten.
- Für jede Person i kann die Veränderungen zwischen der ersten und der zweiten Erhebung anhand der Differenzen der gepaarten Daten

  $$d_i = (x_{i1} - x_{i2})$$

  beschrieben werden
- Die Verteilung dieser Differenzen zeigt die Veränderungen für die ganze Stichprobe.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Sind alle Veränderungen rein zufällig zustande gekommen, so gleichen sich positive und negative Veränderungen in etwa aus: Der Mittelwert der Differenzen

$\bar{x}_d$

wird ungefähr bei Null liegen.
Wir nutzen diese Erkenntnis für die ‚Übersetzung' unserer inhaltlichen Frage in eine statistisch bearbeitbare Frage.
Diese lautet: Stammen die Differenzen der gepaarten Daten aus einer Population, in der diese Differenzen mit dem Mittelwert

$\mu_d = 0$

verteilt sind?

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

3. Prüfgrösse

Die Grösse, die es zu untersuchen gilt, ist somit der Mittelwert der Differenzen der gepaarten Daten.
Prüfgrösse:

$\bar{x}_d$

Wie ist nun diese Prüfgrösse verteilt?

4. Arbeitshypothese und Alternativhypothese

Die Statistiker nennen uns die Verteilung der Prüfgrösse für den Fall, dass die Differenzen der gepaarten Daten tatsächlich aus einer Population stammen, in der diese Differenzen mit

$\mu_d = 0$

verteilt sind, d.h. wenn die sogenannte Arbeitshypothese H_0 gültig ist.
Arbeitshypothese H_0:

$\mu_d = 0$

Zu jeder Arbeitshypothese kann eine Alternativhypothese formuliert werden, die gerichtet oder ungerichtet und spezifisch oder unspezifisch sein kann. Wir wählen als Beispiel die
Alternativhypothese H_1:

$\mu_d > 0$

Diese Alternativhypothese ist **gerichtet** und **unspezifisch**. Sie postuliert, dass die Differenzen der gepaarten Daten aus einer Population stammen, in der diese Differenzen mit dem Mittelwert

$\mu_d > 0$

(gerichtet) verteilt seien. Sie macht indessen keine Angabe, um welchen Betrag

μ_d

von Null abweicht (unspezifisch). Eine solche Alternativhypothese formulieren wir immer, wenn wir uns für eine Veränderung mit einem spezifischen Vorzeichen interessieren, aber nichts über die Grösse der Veränderung postulieren können.

5. Prüfverteilung und Übertretungswahrscheinlichkeit

Prüfverteilung

Sind die Differenzen der gepaarten Daten in der Population normalverteilt und ist H_0 gültig, so ist die Verteilung der Prüfgrösse

$\bar{x}_d$

, d.h. die Prüfverteilung, bekannt. Je nach der Grösse n der Stichprobe müssen zwei Fälle unterschieden werden.

- Grosse Stichproben, d.h. n > 30: Normalverteilung
- Kleine Stichproben, d.h. n # 30: t-Verteilung mit dem Freiheitsgrad df = (n – 1)

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Jetzt kennen wir die Form der Prüfverteilung, es fehlen uns noch die Parameter der Prüfverteilung.
Parameter der Prüfverteilung:

$$\mu_{\bar{x}_d} = 0$$

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}}$$

;

$$\hat{\sigma}_d = s_d * \sqrt{\frac{n}{n-1}}$$

Achtung: Die Schätzformel für

$\hat{\sigma}_{\bar{x}_d}$

basiert auf einer sog. Punktschätzungen der unbekannten Varianz der Differenzen in der Population (

$\hat{\sigma}_d$

) aus der bekannten Varianz der Differenzen in der Stichprobe (s_d). Diese Punktschätzung setzt voraus, dass die Differenzen der gepaarten Daten in der Population normalverteilt sind.
Damit sind die Form und die Parameter der Prüfverteilung bekannt und wir wissen, wie die Prüfgrösse verteilt ist, wenn die Arbeitshypothese H_0 gültig ist.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Mit der Überschreitungswahrscheinlichkeit p kennen wir die Wahrscheinlichkeit, mit der der Ausprägungsgrad der Prüfgrösse

$\bar{x}_d$

- bei Gültigkeit von H_0 - zufällig erreicht oder überschritten wird.
Ist p sehr klein, so ist es sehr unwahrscheinlich, dass H_0 gültig ist. Wir verwerfen H_0 zugunsten von H_1. Ist p hingegen gross, so ist nicht auszuschliessen, dass die Differenzen der gepaarten Daten aus einer Population stammen in der diese Differenzen mit dem Mittelwert

$\mu_d = 0$

verteilt sind. Wir behalten H_0 bei.
Was nun ‚sehr klein' und ‚gross' bedeutet, beurteilen wir wie immer anhand der Signifikanzniveaus.

6. Beurteilung der Übertretungswahrscheinlichkeit; Signifikanzniveau

Die ermittelte Überschreitungswahrscheinlichkeit beurteilen wir an den üblichen Signifikanzniveaus von 5%, 1% und 0,1 %.
Wir wollen uns dies an einigen Beispielen für grosse Stichproben und für den Fall gerichteter und ungerichteter Alternativhypothesen veranschaulichen. Wählen Sie erst eine Alternativhypothese und dann ein Beispiel.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Was bedeutet nun Verwerfung resp. Annahme der Arbeitshypothese H_0 für unsere inhaltliche Fragestellung?
Wird H_0 verworfen, so darf mit der Irrtumswahrscheinlichkeit # p ausgeschlossen werden, dass die Differenzen der gepaarten Daten aus einer Population stammen, in der diese Differenzen mit

$\mu_d = 0$

verteilt sind.

Dies bedeutet, dass bei der Interpretation des Unterschieds in den Mittelwerten der ersten und der zweiten Erhebung der Zufall mit der Irrtumswahrscheinlichkeit kleinergleich_p.gif ausgeschlossen werden kann.
Bei ungerichteter Alternativhypothese H_1 interpretieren wir nur bezüglich des absoluten Betrags des Unterschieds, bei gerichteter Alternativhypothese H_1 auch bezüglich der Richtung des Unterschieds, d.h. auch bezüglich seines Vorzeichens. In Untersuchungen mit Messwiederholung wird die signifikante Veränderung des Mittelwertes in der Regel der Wirkung des Treatments zugeschrieben.
Wird H_0 beibehalten, so ist nicht auszuschliessen, dass die Differenzen der gepaarten Daten aus einer Population stammen, in der diese Differenzen mit

$$\mu_d = 0$$

verteilt sind.
Dies bedeutet, dass bei der Interpretation des Unterschieds in den Mittelwerten der beiden Erhebung der Zufall nicht ausgeschlossen werden kann. Eine Irrtumswahrscheinlichkeit für diese Aussage kann bei unspezifischer Alternativhypothese H1 nicht angegeben werden.
In Untersuchungen mit Messwiederholung ist in diesem Fall keine Wirkung des Treatments nachweisbar.

7. Zusammenfassung zum Lernschritt

Die folgende Zusammenfassung gilt für den Fall, dass eine Stichprobe zweimal untersucht wurde (z. B. Messwiederholung: Erste Erhebung vor, zweite Erhebung nach einem Treatment). Wir interessieren uns für die Signifikanz des Unterschieds in den Mittelwerten der beiden Erhebungen.

Voraussetzung
Das interessierende Merkmal ist intervall- oder proportional skaliert. Die Differenzen der gepaarten Merkmalsausprägungen müssen in der hinter der Stichprobe stehenden Population näherungsweise normalverteilt sein.

Prüfgrösse

$$\bar{x}_d$$

(Mittelwert der Differenzen der gepaarten Daten)

Arbeitshypothese und Alternativhypothese
H_0:

$$\mu_d = 0$$

Beispiele für H_1:

$$\mu_d \neq 0$$

(ungerichtet, unspezifisch)

$$\mu_d > 0$$

(gerichtet, unspezifisch)

Prüfverteilung
Grosse Stichproben, d.h. n > 30: Normalverteilung
Kleine Stichproben, d.h. n # 30: t-Verteilung mit dem Freiheitsgrad df = (n – 1)
Parameter der Prüfverteilung:

- $\mu_{\bar{x}_d} = 0$
- $\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}}$
  ;
  $\hat{\sigma}_d = s_d * \sqrt{\frac{n}{n-1}}$

Zur manuellen Bestimmung der Überschreitungswahrscheinlichkeit muss der Ausprägungsgrad der Prüfgrösse in die z- resp. t-Verteilung transformiert werden.

$$z_{\bar{x}_d} = \frac{(\bar{x}_d - \mu_{\bar{x}_d})}{\hat{\sigma}_{\bar{x}_d}}$$

resp.

$$t_{\bar{x}_d} = \frac{(\bar{x}_d - \mu_{\bar{x}_d})}{\hat{\sigma}_{\bar{x}_d}}$$

Bei ungerichteter Alternativhypothese H_1: Zweiseitige Überschreitungswahrscheinlichkeit.
Bei gerichteter Alternativhypothese H_1: Einseitige Überschreitungswahrscheinlichkeit.

Beurteilung der Überschreitungswahrscheinlichkeit
p > 5%:
H_2 wird beibehalten. Bei unspezifischen Alternativhypothesen kann keine Irrtumswahrscheinlichkeit angegeben werden.
p # 5 %:
H_0 wird mit der Irrtumswahrscheinlichkeit # p zugunsten von H_1 abgelehnt.

Fallbeispiel (1-11)

Inhaltsübersicht

1. Untersuchungsdesign

Nachdem Frau Fischer alle Einwilligungen und notwendigen Zusagen hat, stellt sie ihr Untersuchungsdesign vor:

- Die Kinder der Gruppe A treffen sich am Dienstagnachmittag in der Aula der Schule.
- Alle erhalten eine gleichgrosse Menge "Stimubrain" zum Trinken (Treatment).
- Die Kinder lösen eine Variante des Merkfähigkeitstests, den sie am Wettkampf lösten. Es sind nicht die selben Aufgaben, sie sind aber gleich gestaltet und vom selben Schwierigkeitsgrad.
- Die Resultate werden anhand des t-Tests mit Messwiederholung ausgewertet.
- Der erste Zeitpunkt der Datenerhebung ist der eigentliche Wettkampftag. Der zweite Zeitpunkt der Datenerhebung ist heute, der Tag, an dem der Versuch durchgeführt wird.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]

Auf der nächsten Seite sind die statistisch relevanten Angaben zusammengestellt, die notwendig sind, um das Fallbeispiel zu lösen.

2. Statistisch relevante Daten

Die Rohdaten der Nachuntersuchung mit 'Stimubrain' erhalten Sie hier

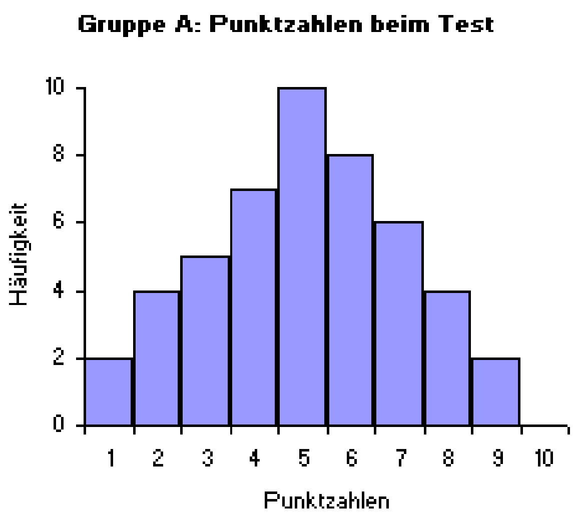
Daten des Tests und Nachtests

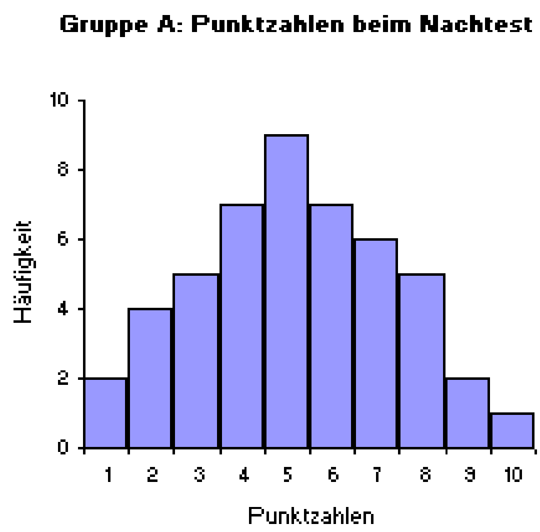
Person	t_{1i}	t_{2i}	$d_i = t_{2i} - t_{2i}$
1	1	2	1
2	1	1	0
3	2	1	-1
4	2	2	0
5	2	2	0
6	2	3	1
7	3	2	1
8	3	3	0
9	3	3	0
10	3	5	2
11	3	5	2
12	4	3	-1
13	4	3	-1
14	4	4	0
15	4	4	0
16	4	5	1
17	4	6	2
18	4	6	2
19	5	4	-1
20	5	4	-1
21	5	4	-1
22	5	5	0
23	5	5	0
24	5	5	0
25	5	5	0
26	5	7	2
27	5	7	2
28	5	7	2
29	6	4	-2
30	6	4	-2
31	6	5	-1
32	6	6	0
33	6	6	0
34	6	7	1
35	6	8	2

Häufigkeit und Histogramm für beide Verteilungen

Häufigkeitsverteilung und Histogramme

Gruppe A: Test		Gruppe A: Nachtest	
Punkte	**Häufigkeit**	**Punkte**	**Häufigkeit**
1	2	1	2
2	4	2	4
3	5	3	5
4	7	4	7
5	10	5	9
6	8	6	7
7	6	7	6
8	4	8	5
9	2	9	2
10	0	10	1
Test: Leistung am Wettkampftag		Nachtest: Leistung in der Nachuntersuchung	





Arithmetischer Mittelwert am Wettkampftag (Test): 5.06
Arithmetischer Mittelwert bei der Nachuntersuchung (Nachtest): 5.21

Bezeichung	Wert	Erklärung
n	48	Anzahl Kinder der Gruppe A
$\bar{x}_d$	0.15	Mittelwert der Leistungsdifferenzen zwischen der zweiten Erhebung (Nachuntersuchung) und der ersten Erhebung (Wettkampftag)
s_d	1.22	Standardabweichung der Verteilung der Leistungsdifferenzen

Im nächsten Schritt werden inhaltliche Fragestellungen vorgestellt.

3. Inhaltliche Fragestellung

Erreichen die Kinder der Gruppe A nach der Einnahme von "Stimubrain" bessere Durchschnittwerte beim Lösen des Merkfähigkeitstests als am Wettkampftag?
Oder:
Sind die Kinder der Gruppe A nach der Einnahme von "Stimubrain" besser beim Lösen des Merkfähigkeitstests als am Wettkampftag?
Oder:
Lösen die Kinder der Gruppe A nach der Einnahme von "Stimubrain" den Merkfähigkeitstest erheblich besser als am Wettkampftag?

Generelle inhaltliche Frage
Darf bei der Interpretation der positiven Differenz zwischen den Leistungen in der Nachuntersuchung und den Leistungen am Wettkampftag der Zufall ausgeschlossen werden?
Im nächsten Schritt wird die Frage so formuliert, dass sie mit statistischen Mitteln bearbeitet werden kann.

4. Statistisch bearbeitbare Fragestellung

Stammen die Differenzen der gepaarten Daten aus einer Population, in der die Differenzen mit dem Mittelwert

$$\mu_d = 0$$

verteilt sind?

Erklärung
Wenn sich die Leistungen zu den beiden Erhebungszeitpunkten Test und Nachtest nur zufällig unterscheiden, dann ergibt sich in der Population aller Leistungsdifferenzen ein Mittelwert von 0, das heisst

$$\mu_d = 0$$

. Implizit heisst das, dass "Stimubrain" hinsichtlich Leistungssteigerung im Merkfähigkeitstest nichts nützt.
Im nächsten Schritt wird die Prüfgrösse bestimmt.

5. Prüfgrösse

Uns interessiert die Differenz zwischen den Leistungen im Nachtest und am Wettkampftag.
Prüfgrösse:

$$\bar{x}_d$$

Nun stellt sich die Frage, wie diese Prüfgrösse verteilt ist, wenn die Differenzen der gepaarten Daten aus einer Population stammen, in der die Differenzen mit

$$\mu_d = 0$$

verteilt sind?
Im nächsten Schritt werden die Arbeitshypothese H_0 und die Alterativhypothese H_1 formuliert.

6. Arbeitshypothese und Alternativhypothese

Arbeitshypothese H_0 :

$$\mu_d = 0$$

Die Arbeitshypothese H_0 postuliert, dass “Stimubrain” nichts bewirkt. Anders ausgedrückt: Die Unterschiede in den Leistungen im Merkfähigkeitstest sind zufällig entstanden.

Alternativhypothese H_1 :

$$\mu_d > 0$$

Die Alternativhypothese H_1 ist gerichtet und unspezifisch. Sie postuliert, dass die Differenzen der gepaarten Daten der Kinder aus einer Population stammen, in der die Differenzen mit Mittelwert

$$\mu_d > 0$$

verteilt sind. Gerichtet ist die Alternative deshalb, weil

$$\mu_d$$

grösser als 0 postuliert wird. H1 macht keine Angaben darüber, wie gross diese Abweichung ist, deshalb ist sie unspezifisch.

Im nächsten Schritt bestimmen wir die Prüfverteilung.

7. Prüfverteilung

Sind die Differenzen der gepaarten Daten in der Population normalverteilt und ist H_0 gültig, so ist die Verteilung der Prüfgrösse

$$\bar{x}_d$$

, also die Prüfverteilung, bekannt: Sie ist mit den Parametern

$$\mu_{\bar{x}_d}$$

und

$$\hat{\sigma}_{\bar{x}_d}$$

normalverteilt, normalverteilt deshalb, weil n > 30.

Parameter der Prüfverteilung:

$$\mu_{\bar{x}_d} = 0$$

mit

$$\hat{\sigma}_d = 1.22 * \sqrt{\frac{48}{47}} = 1.23$$

ergibt sich

$$\hat{\sigma}_{\bar{x}_d} = \frac{1.23}{6.93} = 0.18$$

Wollen Sie die Formeln?

Mit

$$\hat{\sigma}_d = s_d * \sqrt{\frac{n}{n-1}}$$

ergibt sich

$$\hat{\sigma}_{\bar{x}_d} = \frac{\hat{\sigma}_d}{\sqrt{n}}$$

Nun kennen wir die Form und die Parameter der Prüfverteilung unter der Voraussetzung, dass die Arbeitshypothese H_0 gültig ist.
Warum muss vorausgesetzt werden, dass die Differenzen der gepaarten Daten in der Population normalverteilt sind?

Wollen Sie eine Erklärung?

Die Standardabweichung

$$\sigma_d$$

der Differenzen in der Population muss mit einer “Punktschätzung” aus der Standardabweichung s_d der Differenzen in der Stichprobe geschätzt werden. Diese Punktschätzung setzt voraus, dass die Differenzen in der Population normalverteilt sind.

Im nächsten Schritt transformieren wir den Ausprägungsgrad der Prüfgrösse in die z-Verteilung.

8. Prüfgrösse und Übertretungswahrscheinlichkeit

Transformation der Prüfgrösse
Wir transformieren den Ausprägungsgrad der Prüfgrösse

$$\bar{x}_d$$

in die Standardnormalverteilung.

$$z_{\bar{x}_d} = \frac{(\bar{x}_d - \mu_{\bar{x}_d})}{\hat{\sigma}_{\bar{x}_d}}$$

$$z_{\bar{x}_d} = \frac{0.15}{0.18} = 0.83$$

Übertretungswahrscheinlichkeit
Zur Ermittlung der Überschreitungswahrscheinlichkeit der z-transformierten Prüfgrösse

$$z_{\bar{x}_d}$$

benutzen wir die Tabelle A2 (z-Verteilung, Unterschreitungswahrscheinlichkeiten). Wir wählen diese Tabelle, weil H_1 gerichtet formuliert ist und damit die einseitige Überschreitungswahrscheinlichkeit von Interesse ist. Der Ausprägungsgrad von +0.83 wird mit einer Wahrscheinlichkeit von mehr als 20% erreicht oder überschritten.

Beurteilung der Übertretungswahrscheinlichkeit
Die Überschreitungswahrscheinlichkeit ist grösser als 5%, damit wird H_0 beibehalten.
Im nächsten Schritt wird das Ergebnis interpretiert.

9. Interpretation

Die Beibehaltung der Arbeitshypothese H_0 bedeutet, dass bei der Interpretation der Leistungsdifferenzen zwischen den beiden Erhebungen der Zufall nicht ausgeschlossen werden kann.
Das heisst, dass bei den Kindern der Gruppe A durch die Einnahme von "Stimubrain" keine Leistungssteigerung nachweisbar ist, die nicht auch zufällig zustande gekommen sein kann.
Auf der nächsten Seite wird die Datenanalyse mit SPSS vorgestellt.

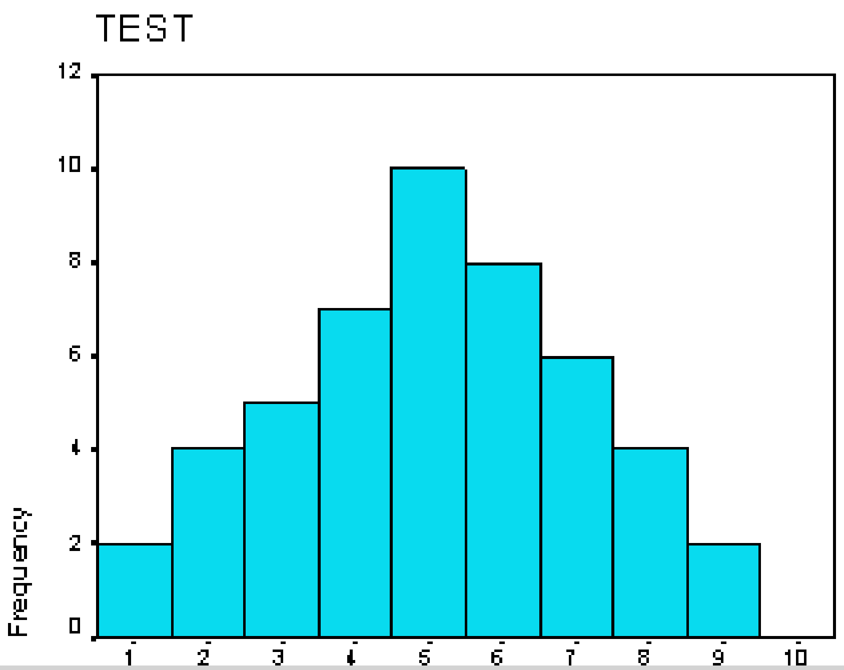
10. SPSS-Ausdruck zum t-Test für abhängige Stichproben

Häufigkeitsverteilung und Histogramm der Ergebnisse der Gruppe A am Wettkampftag Test.

SPSS Befehl:
FREQUENCES VARIABLES = test / HISTOGRAMM.

TEST

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	4.2	4.2	4.2
	2	4	8.3	8.3	12.5
	3	5	10.4	10.4	22.9
	4	7	14.6	14.6	37.5
	5	10	20.8	20.8	58.3
	6	8	16.7	16.7	75.0
	7	6	12.5	12.5	87.5
	8	4	8.3	8.3	95.8
	9	2	4.2	4.2	100.0
	Total	48	100.0	100.0	

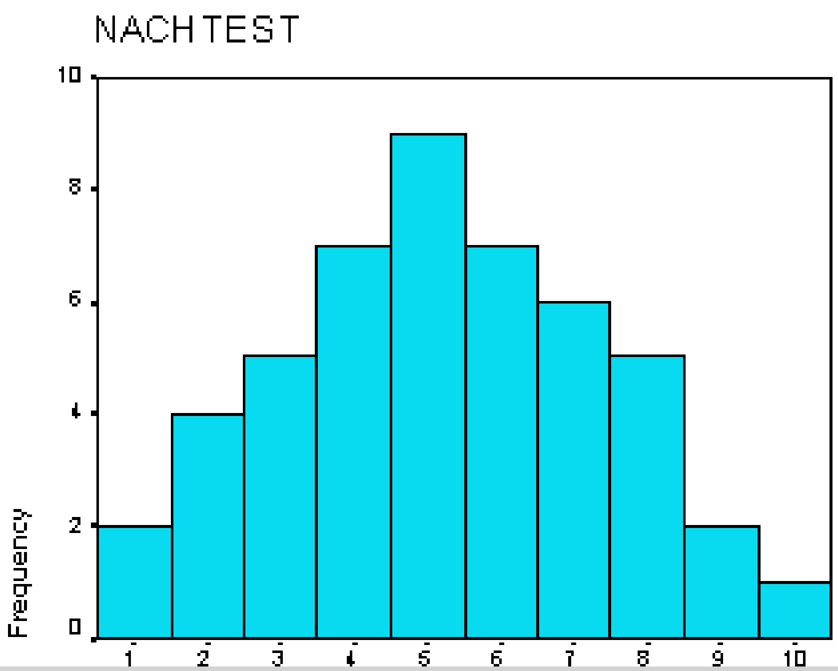


Häufigkeitsverteilung und Histogramm der Ergebnisse der Gruppe A im Nachtest.

SPSS Befehl:
FREQUENCES VARIABLES = nachtest / HISTOGRAMM

NACHTEST

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	4.2	4.2	4.2
	2	4	8.3	8.3	12.5
	3	5	10.4	10.4	22.9
	4	7	14.6	14.6	37.5
	5	9	18.8	18.8	56.3
	6	7	14.6	14.6	70.8
	7	6	12.5	12.5	83.3
	8	5	10.4	10.4	93.8
	9	2	4.2	4.2	97.9
	10	1	2.1	2.1	100.0
	Total	48	100.0	100.0	



t-Test für abhängige Stichproben: Vergleich der Leistungen der Gruppe A in der Nachuntersuchung mit “Stimubrain” (Nachtest) und am Wettkampftag (Test).

SPSS-Befehl
T-TEST /PAIRS = nachtest WITH test.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	NACHTEST	5.21	48	2.19	.32
	TEST	5.06	48	2.04	.29

Paired Samples Test

	Paired Differences			t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean			
Pair 1 NACHTEST - TEST	.15	1.22	.18	.828	47	.412

SPSS nennt den Ausprägungsgrad der Prüfgrösse (Mean), den z-transformierten Ausprägungsgrad (immer als t-Wert bezeichnet) und seine zweiseitige Überschreitungswahrscheinlichkeit (Sig. (2-tailed)). Da die Prüfverteilung (bei grossen Stichproben eine Normalverteilung, bei kleinen Stichproben eine t-Verteilung) symmetrisch ist, entspricht die einseitige Überschreitungswahrscheinlichkeit der Hälfte der zweiseitigen Überschreitungswahrscheinlichkeit.
In unserem Fall ist die einseitige Überschreitungswahrscheinlichkeit grösser als 20%; das Resultat entspricht unserer manuellen Auswertung.

Auf der nächsten Seite gibt Frau Fischer eine Zusammenfassung über ihr methodisches Vorgehen bei der statistischen Bearbeitung der Frage, ob die Gruppe B am Wettkampf gedopt war oder nicht.

11. Frau Fischers Zusammenfassung

Ausgangspunkt des Fallbeispiels war, dass die Schülergruppe B bei einem Wettkampf beim Merkfähigkeitstest bessere Leistungen erzielte als die Gruppe A. Weil bei der Schülergruppe B das Getränk "Stimubrain" gefunden wurde, geriet sie unter den Verdacht, gedopt gewesen zu sein. Auf Anfrage der Schulpflege erklärte ich mich bereit, dem Verdacht mit statistischen Mitteln nachzugehen. Ich wählte folgendes Vorgehen:

1. Ich prüfte, ob sich das Merkmal "Leistung im Merkfähigkeitstest" in beiden Gruppen in etwa gleich um die Leistungsmittelwerte verteilt. Anhand des **F-Tests** konnte ich zeigen, dass das Merkmal in beiden Gruppen ähnliche Varianzen aufweist: es liegen homogene Varianzen vor. In der Streuung der Resultate konnte ich keine Unregelmässigkeiten feststellen.
2. Nun wurde mit dem **t-Test für unabhängige Stichproben** geprüft, ob sich die beiden Stichproben hinsichtlich des Mittelwertes signifikant unterscheiden. In der Tat, die Mittelwerte der Kindergruppen unterschieden sich erheblich und signifikant. Achtung: Das heisst nicht, dass die Kindergruppe B gedopt war! Für den Unterschied können viele Gründe verantwortlich sein.
3. Im Rahmen einer "Vorher-Nachher-Untersuchung" mit der Gruppe A konnte anhand des **t-Tests für abhängige Stichproben** geprüft werden, ob sich die Leistungsfähigkeit der Kinder der Gruppe A unter dem Einfluss von “Stimubrain” signifikant verändert. Der **t-Test für abhängige Stichproben** ergab, dass die geringe Leistungssteigerung, die sich bei der Gruppe A durch die Einnahme von "Stimubrain" ergab, auch zufällig entstanden sein kann.

Schlussfolgerung

"Stimubrain" hat keinen nachweisbaren Einfluss auf die Leistung in der Merkfähigkeit. Somit ist die bessere Leistung der Gruppe B am Wettkampftag nicht auf den Genuss von "Stimubrain" zurückzuführen.
Kurz gesagt: Die Gruppe B war nicht gedopt.
Somit sind Sie am Ende des Lernschrittes **t-Test für abhängige Stichproben** angelangt.